

Use of bypassing in a pipelined instruction processor

The invention relates to an instruction processing device with a pipelined functional unit.

Data dependency imposes significant restrictions on the throughput of instruction processors. Instruction processors execute successive operations that require input
5 operands and produce results. Operands are generally stored in a register file, from which they are retrieved using operand addresses from a command in an instruction. The result is stored in the register file using a result address of the command. When its operands are retrieved from the register file the command cannot be executed until the operands have been stored in the register file by preceding instructions. Thus a minimum delay between the
10 commands is needed. This reduces efficiency of the processor. In VLIW processors, for example, no-operations may have to be scheduled for functional units in some instruction cycles because insufficient operands are available.

US patent No. 5,805,852 describes how a VLIW processor can be made more efficient by means of a bypass between a pipeline stage of a functional unit in which results
15 are produced and a pipeline stage wherein operands are used. The bypass makes the result available as operand for a subsequent instruction without the delay necessary for storing the result in the register file and retrieving it as an operand from the register file.

During pipelined operation, a functional unit first generates the result in an execution stage of a pipeline, and stores the result in a pipeline register behind the execution
20 stage. Subsequently the functional unit hands on the result through the pipeline until it has been stored in the register file. When a new command enters the pipeline of one of the functional units, its operand addresses are compared with the addresses of results that are still in the pipelines of respective ones of the functional units. When a match occurs the operand is taken from the pipeline stage of the relevant functional unit rather than from the register
25 file.

In recent years the size of register files has tended to increase. Large register files have the advantage that they speed up execution because it is less frequently necessary to wait until a register is available for reuse or to spill operands to memory. The price of the

larger register files has been an increase in power consumption. The register file now often is the major power consumer in a VLIW processor.

Amongst others, it is an object of the invention to reduce the power consumption due to register files in an instruction processing device.

5 The instruction processing device according to the invention is set forth in Claim 1. The invention is based on the observation that power can be saved by not writing results to a register file when they do not need to be retrieved from the register file, because they are used only via a bypass path. Prior to execution of instructions, for example during compilations of the instructions, it can be determined whether it can be guaranteed that a
10 result of an operation will be not used other than via a bypass. If so, bypassing of the result suffices and it is not necessary to store the result in the register file. By disabling storage into the register file in this case, power consumption is reduced.

 In an embodiment the instruction processing device contains a plurality of bypass registers that are selectively addressable with a register address from the command,
15 for selecting a register for storing the result and/or for retrieving operand data. Thus, it is made possible to avoid writing to the register file more often. Typically there are far fewer bypass registers than registers in the register file, so as not to slow down the instruction cycle duration. The result is written to a bypass register at a pipeline stage before it is written to the register file.

20 In an embodiment writing of the result is disabled by suppressing a supply of clock signals to circuitry for writing the result into the register file. When the register file has a number of write ports writing is disabled at selected write ports, selected under control of the instructions.

 Preferably, bypassing is controlled by a bypass control unit that compares
25 result register addresses from the commands with operand register address of later commands causing substitution of a result from a bypass path in case of a match of the addresses. Thus, no special addresses are needed for bypassed results. In another embodiment the instructions contain addresses to select between different bypass registers.

 In an embodiment a chain of registers is provided for supplying bypass
30 operand data. Results shift through bypass registers in the chain in successive instruction cycles. The chain extends further than necessary for writing the result into the second register unit. This makes it possible more often to avoid power consumption for writing results to the register file. Bypass data from the registers in the chain may be selected by comparing

operand addresses with result register addresses, or using explicit register selection information from the instructions. The latter simplifies bypassing control circuitry.

The invention is advantageously applied to processors such as VLIW processors, which contain a plurality of functional units that operate in parallel. Such processors require increasingly larger register files since more functional units operate in parallel. By suppressing writing to the register files considerable power consumption is saved. Preferably groups of bypass registers are provided, each for storing results from a respective one of the functional units only, the registers of all groups being addressable from each command for retrieving operands.

The invention also relates to a method of compiling programs, in which the conditions for suppressing writing to the register file are detected, after which information is added to the instructions to suppress such writing. Detection involves testing whether results of instructions can be passed via a bypass path (this is mainly a matter of being used sufficiently soon after production) and whether it can be guaranteed that these results will not be used later (e.g. by scanning the instructions to detect whether the result is not used again in any later reachable instruction before the register that contains the result is overwritten or before the end of the program). A computer program for executing such a method may be passed on any computer program product, such as a magnetic or optical disc, a semiconductor memory module, a download signal etc.

These and other objects and advantageous aspects of the invention will be described using the following figures.

Figure 1 shows a pipelined processor

Figure 2 shows part of a register file

Figure 3 shows part of a pipelined processor

Figure 4 shows part of a further pipelined processor

Figure 5 shows part of a further pipelined processor

Figure 1 shows an example of a simplified pipelined VLIW processor. The processor contains an instruction memory 10, a program counter 10a, an instruction register 11, execution units 12, a register file 14 and a bypass control unit 16. By way of example, two execution units 12 are shown in parallel, but in practice more execution units may be

used. Each execution unit may contain a group of functional units (not shown), or be a functional unit by itself. Instruction register 11 has outputs for a plurality of commands from an instruction, each command for a respective one of execution units 12. Each command contains a part for an opcode, a part for operand register addresses and a part for a result register address. The outputs of instruction register 11 for the operand register address parts of the commands are coupled to read ports of register file 14 and to operand register address inputs of bypass control unit 16. Usually, each command contains two operand addresses, but for the sake of clarity connections for only one operand (address) are shown. More than two operands are also possible for operations such as the multiply-accumulate.

The processor is divided into successive pipeline stages, which are separated by means of pipeline registers. For each execution unit 12 the processor contains first stage pipeline registers 120, 122, 124 and a multiplexer 123 between instruction register 11 and the execution unit 12. A first one of the first stage pipeline registers 124 stores the opcode part of the command for the execution unit 12. A second one of the first stage pipeline registers 122 stores operands of the command for the execution unit 12. A third one of the first stage pipeline register 120 stores the result address of the command for the execution unit 12 and write control information. The first one of the first stage pipeline registers 124 has an input coupled to the outputs of instruction register 11 for the opcode part of the command for the execution unit 12. The third one of the first stage pipeline registers 120 has an input coupled to the outputs of instruction register 11 for the result address part of the command for the execution unit 12.

The second one of the first stage pipeline registers 122 has an input coupled, via multiplexer 123, to the read port of register file 14 to which the operand address parts of the command for the execution unit 12 are supplied. In principle, there will be a respective multiplexer 123 and a respective second one of the first stage pipeline registers 122 with similar connections for each of respective one of the operands of the command for the execution unit, but for the sake of clarity only one multiplexer 123 and second one of the first stage pipeline registers 122 is shown.

Second stage pipeline registers 126, 128 are included behind execution units 12. A first one of the second stage pipeline registers 126 is coupled to the third one of the first stage pipeline registers 124, for receiving the result register address parts of commands and the write control information. A second one of the second stage pipeline registers 128 is coupled to a result output of execution unit 12. The first and second ones of the second stage

pipeline registers 126, 128 are coupled to write ports of register file 14, for supplying results and corresponding result register addresses, as well as the write control information.

5 Multiplexers 123 each have an input coupled for receiving an addressed operand from the read ports of register file 14, and for receiving bypass operands from the second ones of the second stage bypass registers 128 via bypass paths 15. The third one of the first stage registers 124 and the first one of the second stage registers 126 pass operand register addresses and result addresses to bypass control unit 16 respectively. Bypass control unit 16 controls multiplexers 123 to determine which of their inputs is coupled to the second one of the first stage pipeline registers 122.

10 In operation program counter 10a supplies a series of instruction addresses to instruction memory 10. In response to each instruction address instruction memory supplies a respective instruction to instruction register 11. Each instruction contains commands for a plurality of execution units 12. The commands may contain operand register addresses of the operand or operands of the command. The operand register addresses are supplied to read
15 ports of register file 14. In response register file 14 supplies the addressed operands from the read ports. Normally, the operands are supplied to the relevant execution units 12 with the (optionally decoded) opcode part of the corresponding command. The execution units 12 execute the commands using the operand and produce a result. For an "ADD" command for example, two operands are used and their sum is produced as a result. The result is supplied
20 to a write port of register file 14, together with a result register address from the command.

Command execution is pipelined. This means that a new instruction starts in substantially every instruction cycle and that successive steps of execution of an instruction are executed in successive instruction cycles. For example in a first instruction cycle the instruction memory are addressed, in a second instruction cycle the operands are fetched, in a
25 third instruction cycle the command is executed proper by the execution unit 12 and in a subsequent instruction cycle the result is written to register file 14. Thus, execution of an instruction takes a number of instruction cycles. The respective parts of the processor that execute the instruction in respective successive instruction cycles are called pipeline stages. In a particular instruction cycle the different pipeline stages process different instructions in
30 different stages of execution. The different pipeline stages are separated by pipeline registers to separate the information of the different instructions under process. Instruction cycles are indicated by a clock (not shown) which control takeover of information from a preceding pipeline stage into the pipeline register at the end of each instruction cycle.

Pipelined processors are known per se. It will be appreciated that the figure shows only one simple embodiment. Without deviating from the invention many alternative pipelined architectures may be used, in which different tasks are distributed differently over different pipeline stages and that additional pipeline stages may be added.

5 Bypass paths 15 serve to reduce the need to wait with execution of an instruction until its operands have been stored in register file 14. If a first instruction produces a result that is used as operand by a second instruction, the result can be passed to execution unit 12 via bypass paths 15 before the result has been written to register file 14. Bypass control unit 16 compares the operand register address of the second instruction with
10 the result register address of the first instruction. When a match occurs bypass control unit 16 controls multiplexer 123 to pass the result from pipeline register 122 instead of a result from register file 14. This makes it possible to execute the second instruction sooner.

 Bypassing itself is known per se. It should be appreciated that many forms of bypassing exist and that the form shown in figure 1 is but one example, to which the
15 invention is not limited. For example, the results used for the bypass may come from different pipeline stages than the one shown (e.g. directly from execution unit 12 to first stage pipeline register 122, or from a later stage (not shown)), or from more than one pipeline stage. Correspondingly, result register addresses may come from different stages than the one shown. Such addresses or comparison results may be stored in bypass control unit 16 to
20 pipeline comparison as well. In addition a code may be present in the instructions to indicate whether bypassing should be used or not. This makes it possible to avoid bypassing when an instruction should use an "old" value from register file 14.

 According to the invention write back of the result to register file 14 can be disabled under control of the instructions. This will be done when the result is used only via
25 the bypass path 15 and not from register file 14. During compilation of a program of instructions a guaranteed last instruction of the program is determined that uses a result for each execution path that can be followed (if applicable via branch instructions). It will be appreciated that some execution paths may in fact never be reached during execution, but if this cannot be proven at compile time for some path, the last instructions that uses the result
30 if that path is followed should be determined. If any last instruction that has been determined is scheduled to be executed so shortly after the instruction that produces the result that it can receive the result via bypassing, the compiler adds disable information to the instruction that produces the result. This indicates that the instruction need not be stored in register file 14.

The disable information is passed down the pipeline with the result register address to register file 14 to disable storage of the result in register file 14.

Figure 2 shows an example of a multiport register file with multiport registers 20 that have plural read ports 28. Data supply circuits 22 and addressing circuits 24 are coupled to multiport registers 20 for supplying result data and register selection signals respectively. Data supply circuits 22 have inputs RES1, RES2 for results. Addressing circuits 24 have inputs ADDR1, ADDR2 for result register addresses. Clock enable circuits 26 are coupled between a clock input CLK and clock terminals of data supply circuits 22 and addressing circuits 24. Clock enable circuits 26 are used to disable supply of clock signals to data supply circuits 22 and addressing circuits 24 under control of disable inputs DIS1, DIS2.

In operation clock pulses from clock enable circuits 26 cause data supply circuits 22 and addressing circuits 24 to drive supply of new data and selection signals to multiport registers 20. Information from the instructions is supplied to disable inputs DIS1, DIS2 for selectively disabling supply of clock signals for using data and addresses from those ports RES1, ADDR1, RES2, ADDR2 from which results are supplied for which it has been determined that no further use will be made of the results in the program. As a result power consumption by the data supply circuits 22 and addressing circuits 24 is reduced.

It will be appreciated that multiport register files are known per se. Many alternative architectures may be used without deviating from the invention. According to the invention the register file is arranged to disable selected parts of the register file. Instead of disabling both data supply circuits 22, and addressing circuits only one of these circuits may be disabled and/or any other circuit that consumes power when driven to update register content.

Figure 1 shows that the disable information for a result is passed along with the result register address of a command, e.g. in the form of an additional bit from the command. This requires a minimal amount of modification of the processor to disable writing of the result. However, it will be appreciated that many alternative solutions exist for indicating selective disabling. For example, the information may be contained at the level of the instruction rather than at the level of individual commands, by encoding for example the position of one or more commands in the instruction for which disabling is allowed. In fact, the disable information need not even be contained in the same instruction as the relevant command. Instead it may be contained in an earlier or later instruction that is known to be executed at a defined pipeline delay with respect to the command. Similarly, the information to disable may be contained in the opcode of a command, in which it may be decoded

anywhere in the pipeline for ultimate supply to the disable input of a write port of register file 14. In yet another embodiment the information to control disabling may be supplied from an operand register. In this embodiment, data from an addressed operand register is passed along the pipeline stages to register file 14 to control disabling.

5 Although only a single result register 126 has been shown for each execution unit it will be appreciated that the invention is not limited to such a single register.

Figure 3 shows part of a processor with a plurality of result registers 32. Two operand inputs are shown for execution unit 12. The result output of execution unit 12 is coupled to a result register 128 in parallel to a number of further result registers 32. Result
10 register 128 is coupled to circuitry further down the pipe-line (not shown). A register select unit 30 has an input coupled to a first stage result register address register 124 and select outputs coupled to select inputs of further result registers 32. The outputs of further result registers 128 are coupled to inputs of an operand selection unit 34, which in turn has outputs coupled to multiplexers 123.

15 In operation the processor stores selected results in selected ones of the further registers 32. First stage result register address register 120 receives the result register address and the disable information for disabling writing to register file 14. Register select unit 30 is enabled when the disable information indicates that writing to register file 14 will be disabled. In this case, register select unit 30 uses part of the result register address to select
20 one of the further result registers 32. The result from execution unit 12 is written into the selected further result register 32. Subsequently, when an operand address is received that indicates that a result from a further result register 32 should be used as operand, bypass control unit 16 signals operand selection unit 34 to select the output of one of further result registers 32 on the basis of the operand address. The result is then passed to execution unit 12
25 as operand from operand selection unit 34 via multiplexer 123.

It should be appreciated that the number of further result registers 32 is much smaller than the number of registers in register file 14. This makes it possible to include these further result registers 32 at the end of the pipe-line stage that contains execution unit 12 without a large time penalty. Thus, a limited number of results can be stored at an earlier
30 pipeline stage than the pipeline stage in which results are normally stored in register file 14. These results can be made available as operands from further result registers 32 well before they could have been made available from register file 14.

It should also be appreciated that figure 3 merely illustrates an example of how further result registers 32 may be used. Many alternatives are possible. For example, further

result registers 32 might be included in a subsequent pipeline stage, i.e. following result register 128, or in more than one pipeline stage. The former may be the case for example when inclusion of the further result registers at the executions stage would require an extension of the duration of the instruction cycle. When different execution units 12 are used in parallel, different sets of further result registers 32 may even be included at different pipeline stages. As another example, selection of the result registers 32 might be implemented differently. For example, a dedicated address part might be included in instructions to select a further result register, or on the contrary register address values that do not address physical registers in register file 14 might be used to select further registers 32. As another example a mode register might be used to select between using and not using further result registers 32. Also selection information may be part of the opcode of commands, instead of coming from the result register address field.

Similarly, various solutions are possible for determining when a result from a further result register 32 should be used instead of data from register file 14. In one embodiment operand addresses contain a special indication to indicate that a result from a further result register 32 should be used.

Figure 4 shows an alternative embodiment wherein address matching is used. For each further result register 32 a respective register address register 40 is provided that stores the result register address for the result stored in further result register 32. Bypass control unit compares operand register addresses with the result register addresses from register address registers. When a match occurs bypass control unit 16 signals operand selection unit 34 and multiplexers 123 (shown in Figure 3) to substitute a result from further result registers 32 for operand data from register file 14. Preferably, bypass control unit 16 also compares result register addresses from the pipeline register 124 with stored result register addresses from register address register 40 and resets a register address registers 40 when it contains the register address of specified in a command. Thus it is ensured that a result from register file 14 will be used subsequently for that register address (unless of course a new result is stored in further result registers 32 subsequently). It will be appreciated that further result registers 32 with register address registers 40 form an embodiment of a simple associative memory. Other types of associative memories may be used as an alternative.

Although it has been assumed in the preceding that the instructions contain information to indicate which of the further result registers 32 should be used to store a particular result, it should be appreciated that instead it may suffice to indicate merely that

the particular result and its result register address should be stored in one of the further result registers 32 and its corresponding further result register address register 40, and optionally for how many instruction cycles the result should be stored, or how many times it will be read. In this case, an automatic register assignment unit may be provided to assign the result to any free one of the further result registers 32 (a further register may be recorded as free after a predetermined number of instruction cycles (or the optionally programmed number) has passed, or after the result has been read (or read the programmed number of times)). Although useful, further result register address register 40 is not strictly necessary for this: at compile time it is possible to predict which further result register 32 will be used and a selection address for that further result register 32 may be included in the operand specification.

Although only one execution unit 12 was shown in figures 3 and 4 for the sake of clarity, it will be understood that in practice a plurality of execution units 12 may be used in parallel to execute different commands from an instruction. In this case, each execution unit 12 may be provided with its own set of further result registers 32. Bypass control unit 16 controls whether and from which set results are substituted for operand data from register file 14. As an alternative, a shared set of further result registers 32 may be used. In this case the instructions should contain information to indicate which, if any, of the execution units 12 should write to which further result register 32 (and, if necessary to its corresponding result register address register 40).

It will be appreciated that writing into further result registers can make it superfluous to write into register file 14. By including information in the instructions to selectively write results into further result registers 32 and to disable writing of a result into register file 14 overall power consumption can be reduced. Of course, writing into further result registers also consumes power, but because the number of these further results registers 32 is smaller than that in register file 14 less power is consumed. The size of register file 14 can therefore be extended without a severe power consumption penalty. As shown in figure 3, the same signal is used to signal disabling of writing into register file 14 and to enable writing into further result registers 32. This helps to reduce instruction size. As a result, results are written either in further result registers 32 or in register file 14. However, it is also possible to used independently settable control information for disabling writing into register file 14 and for selecting further result registers 32. Thus, under program control a bypass of a result from further result registers can be selectively combined with either or not longer term storing in register file 14, as needed for a particular program.

Figure 5 shows part of another embodiment of a processor. Compared to the embodiment of figure 1 in the embodiment of figure 5 a number of further result register address registers 50 and further result registers 52 have been added downstream from result register address registers 126 and result registers 128. Further result register address registers 50 and further result registers 52 have outputs coupled to bypass control unit 16. In operation results of execution of commands and result register addresses of these results are passed down the pipeline, even after they the instruction cycle from which they could be retrieved from register file 14 if writing is enabled. Bypass control unit 16 controls whether results from further result register address registers 50 and further result registers 52 are used or from register file 14. This may be done on the basis of register operand/result address comparison or using explicit information from the instruction. Use of further result register address registers 50 and further result registers 52 makes it possible to disable writing of a result into register file 14 when that result is needed only within a predetermined number of instruction cycles after it has been generated. When this is the case, information is added to the instructions to disable writing of the result into register file 14. Thus, power consumption for writing into register file 14 can be saved.

In an embodiment storing of a register address and a result in further result register address registers 50 and further result registers 52 is disabled if the result is stored in register file 14. Thus, power is saved when storing in register file 14 is not disabled. In this embodiment additional registers for disable information are included in parallel with further result register address registers 50. The disable information is coupled to further result register address registers 50 and further result registers 52. The disable information is used to disable updating of the content of further result register address registers 50 and further result registers 52 when the disable information indicates that writing of the result into register file 14 has not been disabled. In this case, the disable information may also be used to indicate to bypass control unit that the operand data should be retrieved from register file 14 instead of further result register address registers 50 and further result registers 52.

It will be understood that any number of one or more further result register address registers 50 and further result registers 52, not only two, may be included in series, permitting results to be bypassed as operands. It will also be understood that additional result register address registers 50 and result registers 52 may be present in the pipeline between the result register address register 126 and the result registers 128 that follow execution unit 12 on one hand and further result register address registers 50 and further result registers 52 on

the other hand when there are additional pipeline stages behind execution unit. Bypass control unit 16 uses results from these stages for bypass as well.